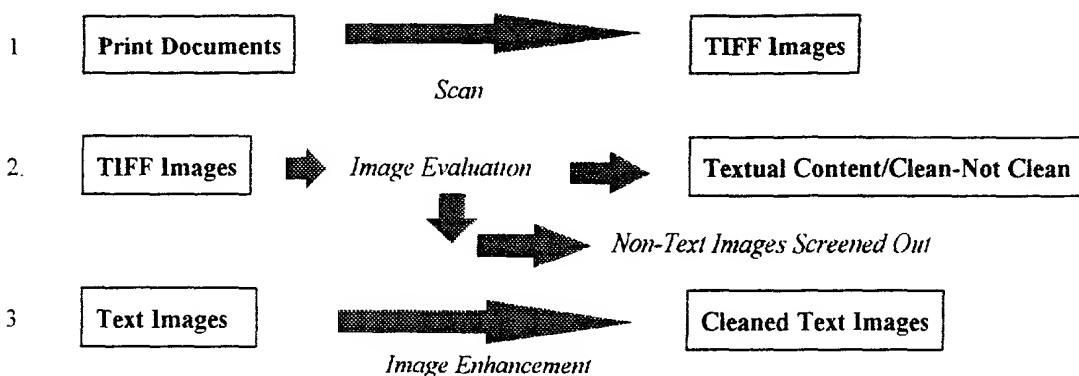


Figure 1

Image-to-Data Processing System:

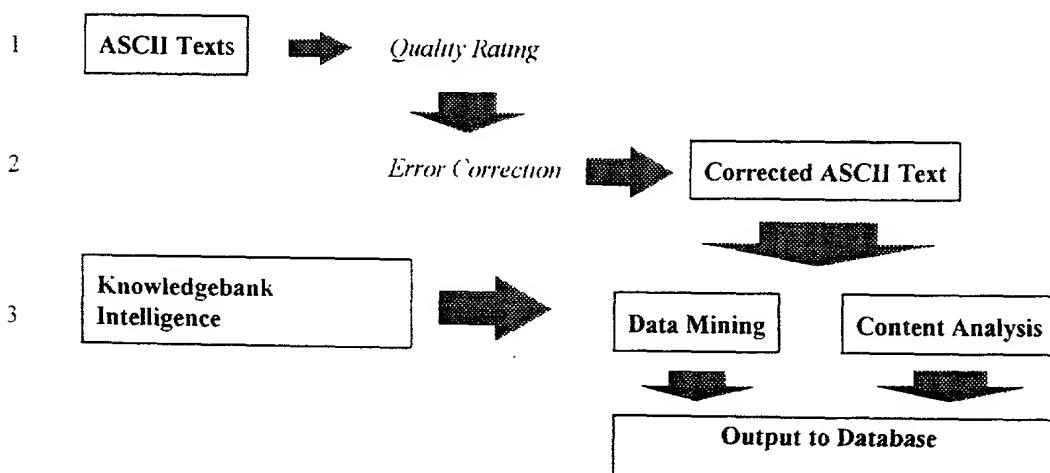
Pre-OCR Processes



Optical Character Recognition: Image-to-Text Conversion



Post-OCR Processes: Text-to-Data Conversion



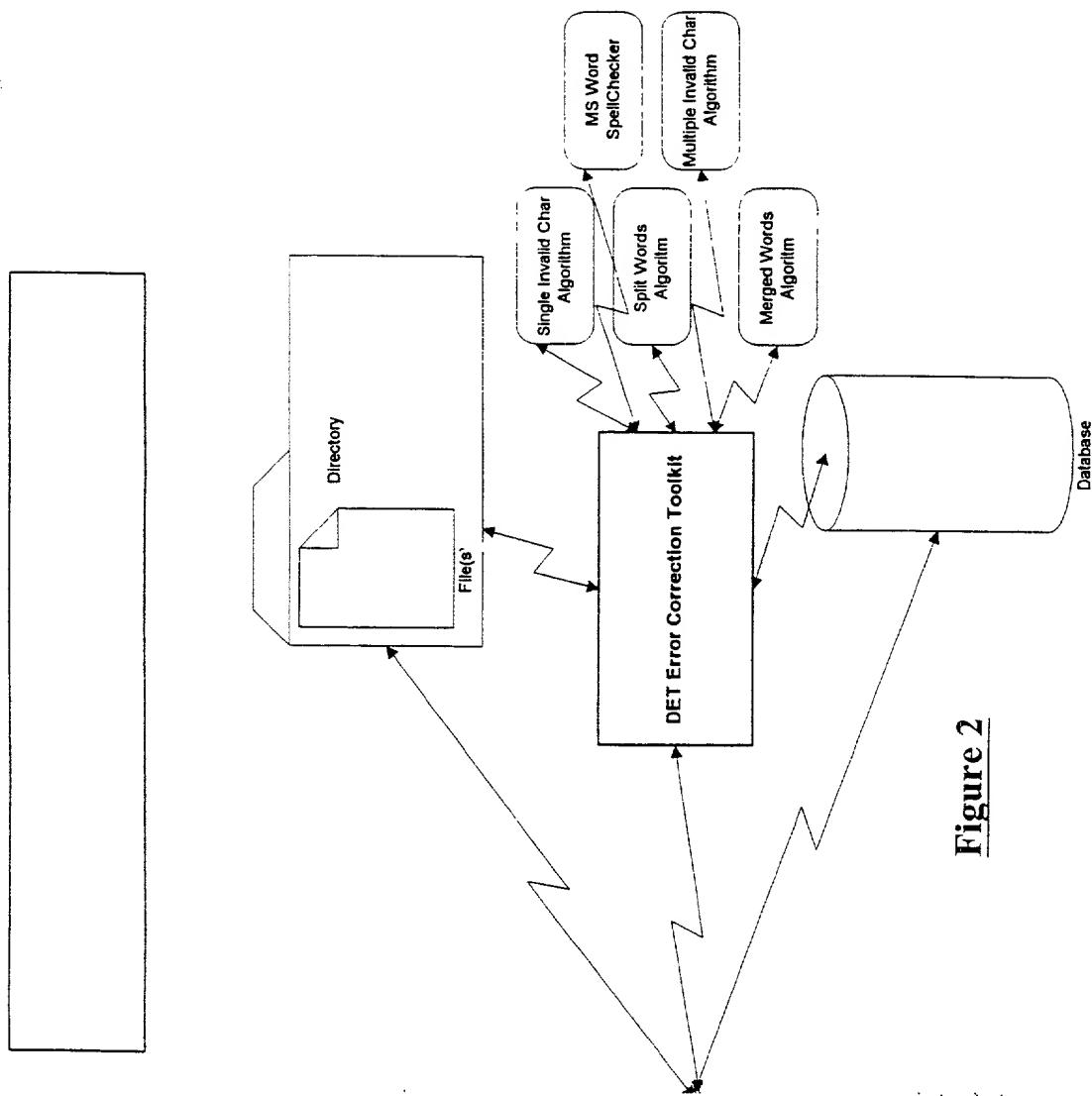
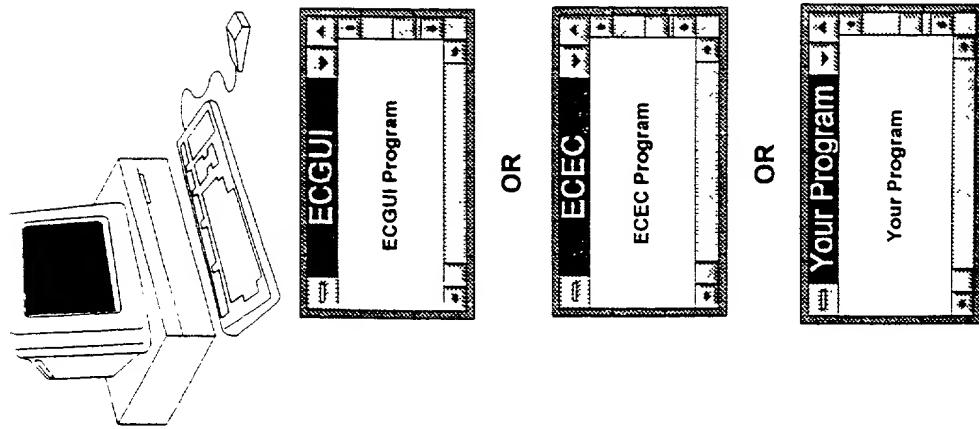


Figure 2



Computer-Aided Error Correction

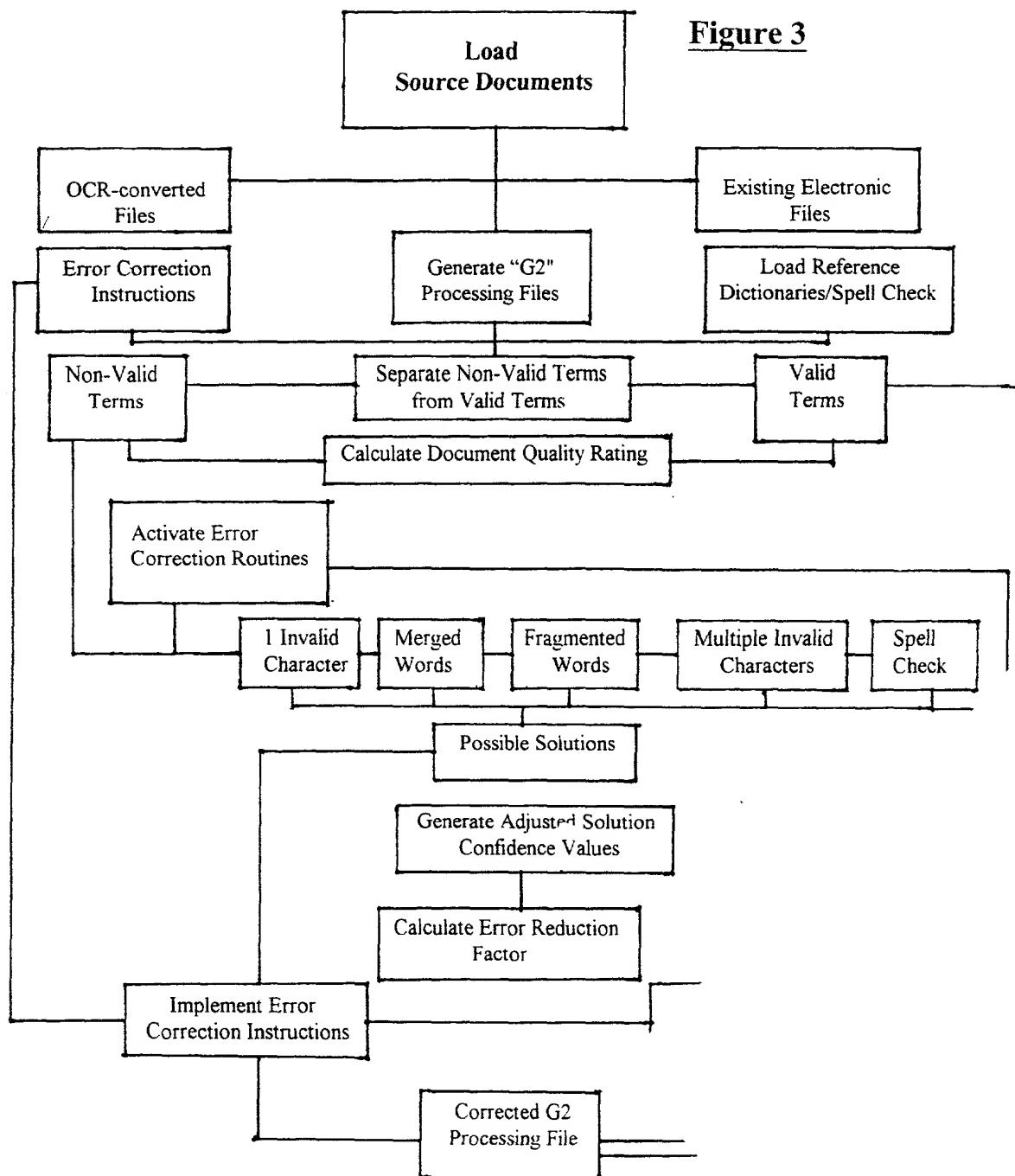


Figure 4

Diagram: Screening

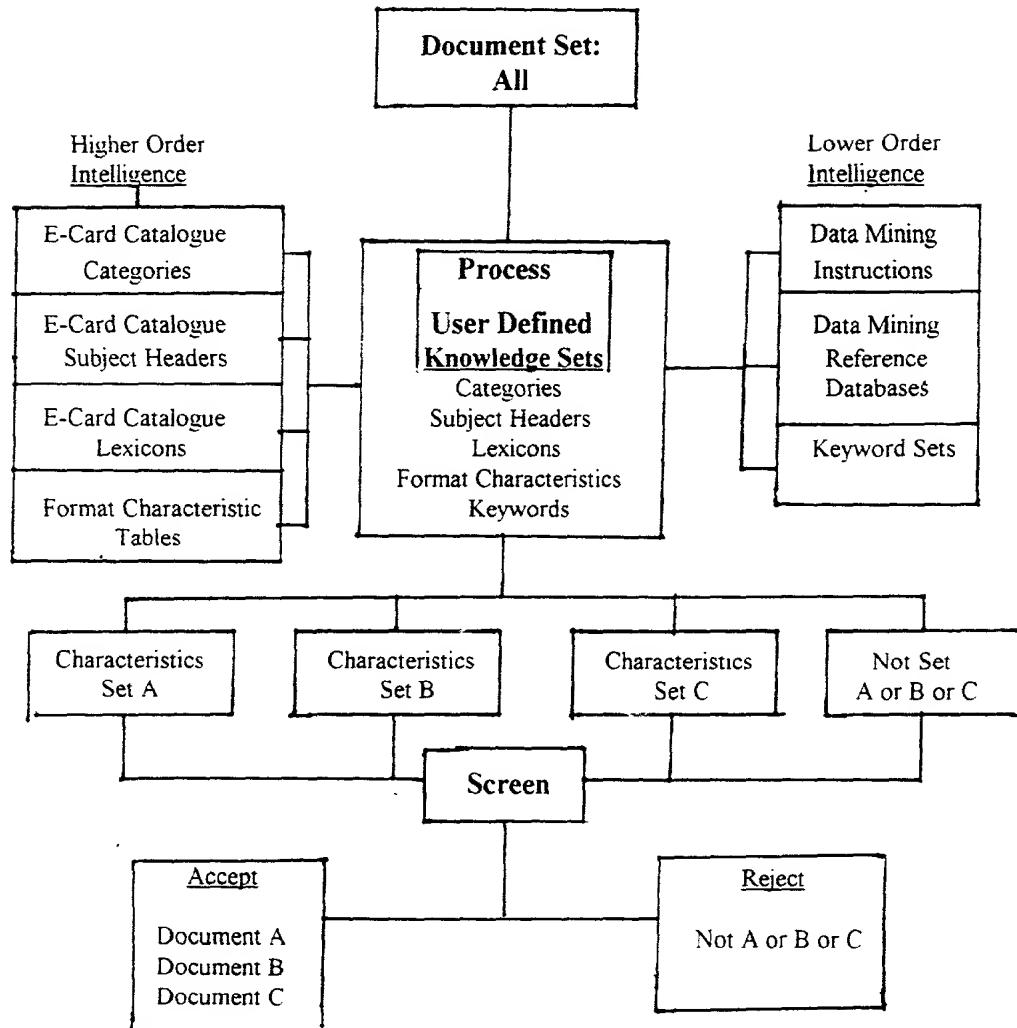


Figure 5

Diagram: Sorting

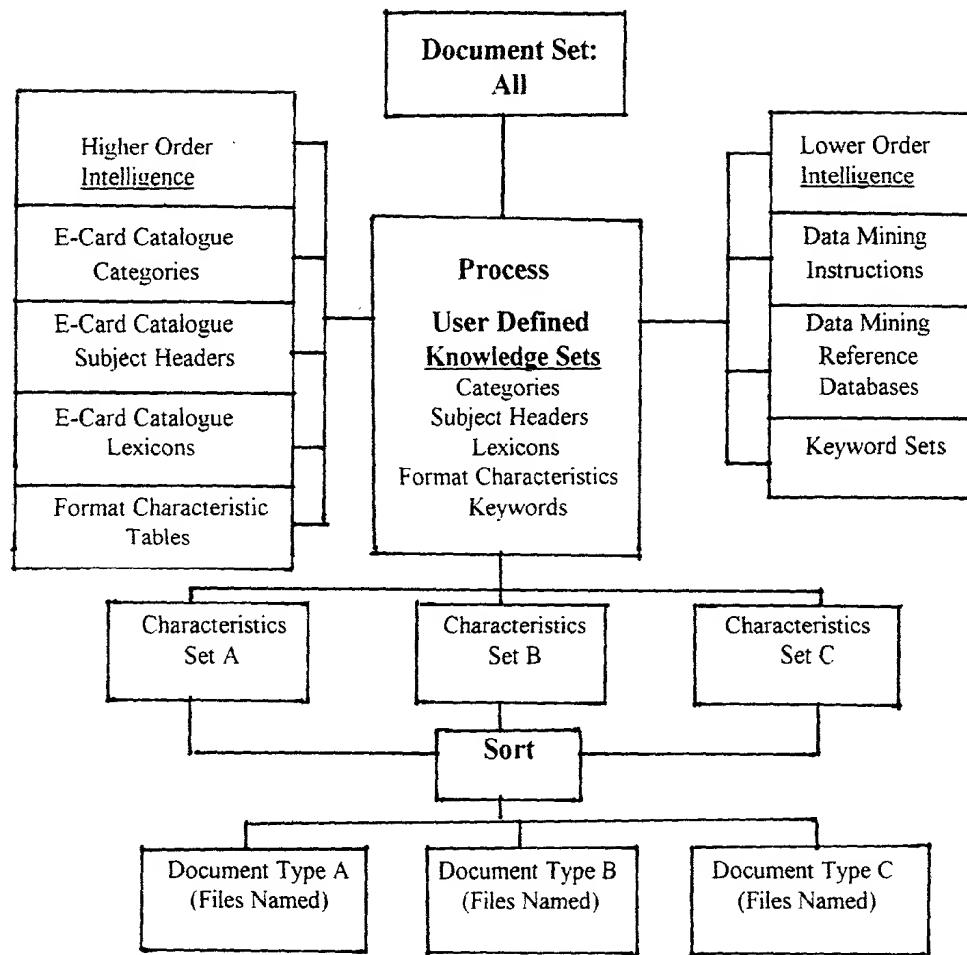


Figure 6

Diagram: Characterizing

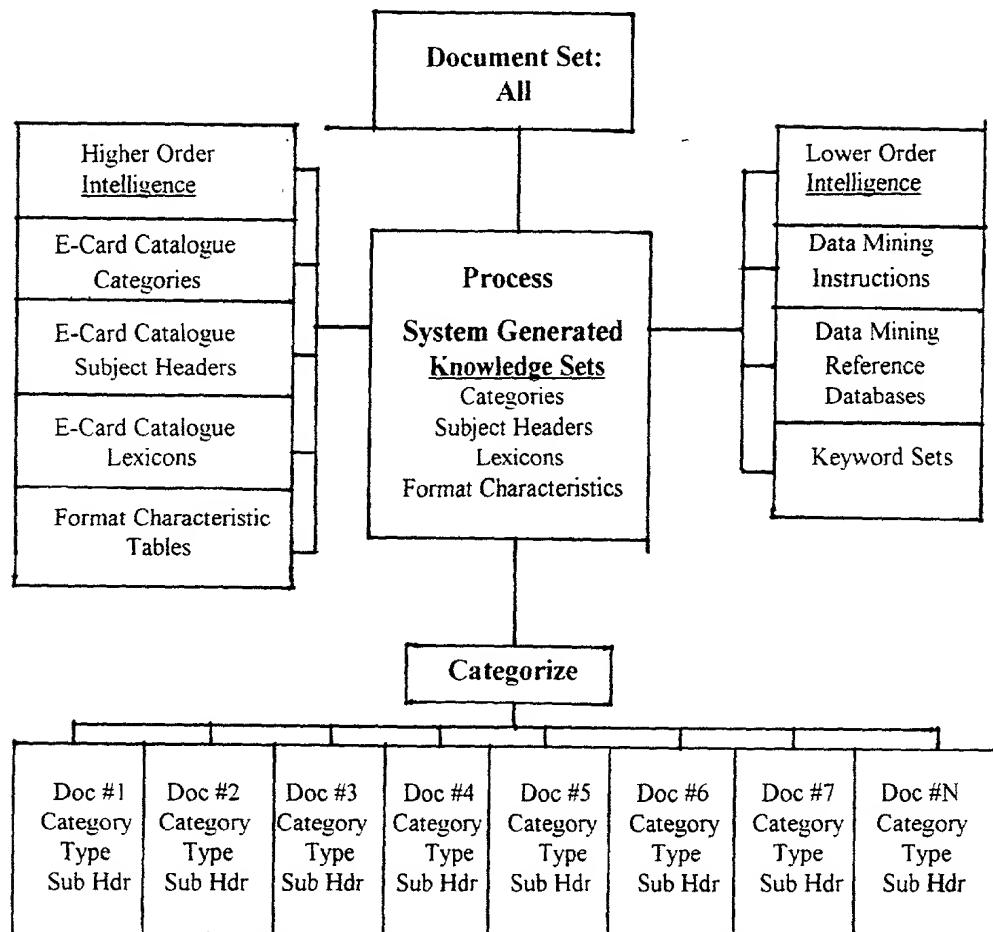


Figure 7

Diagram: Abstracting

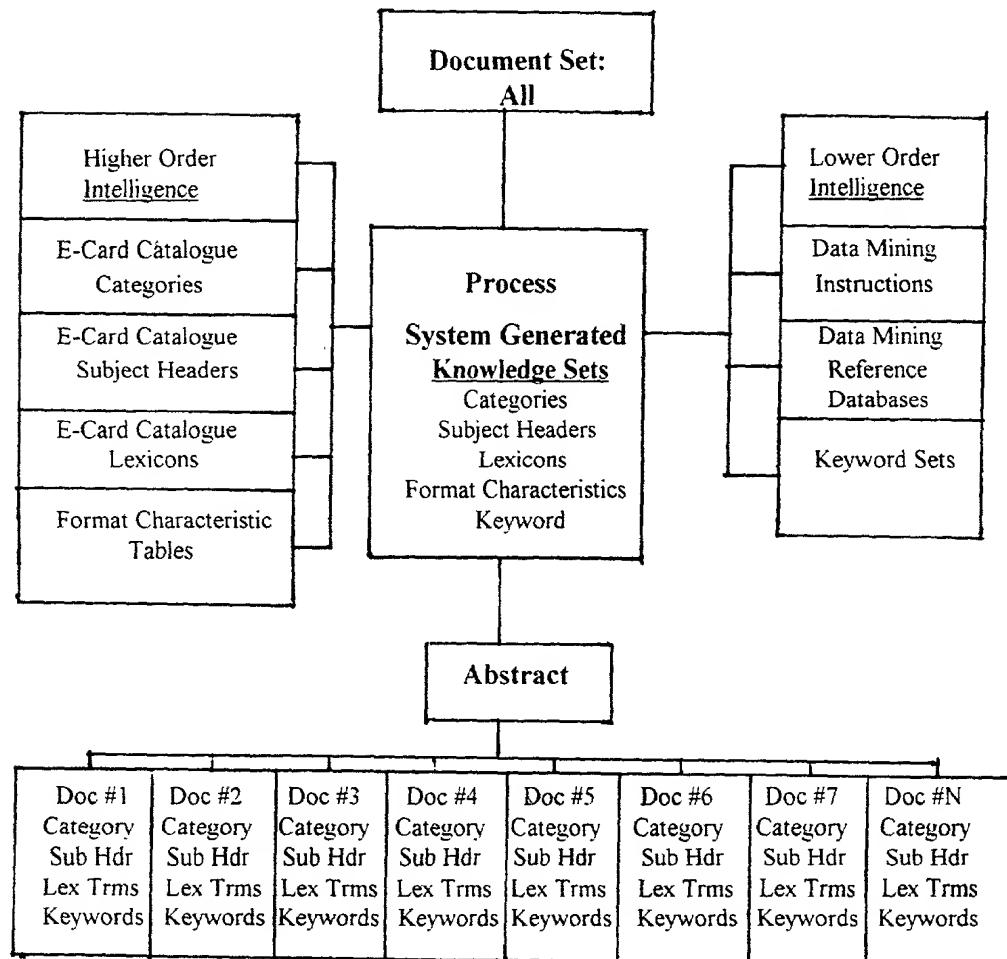


Figure 8

Diagram: Profiling

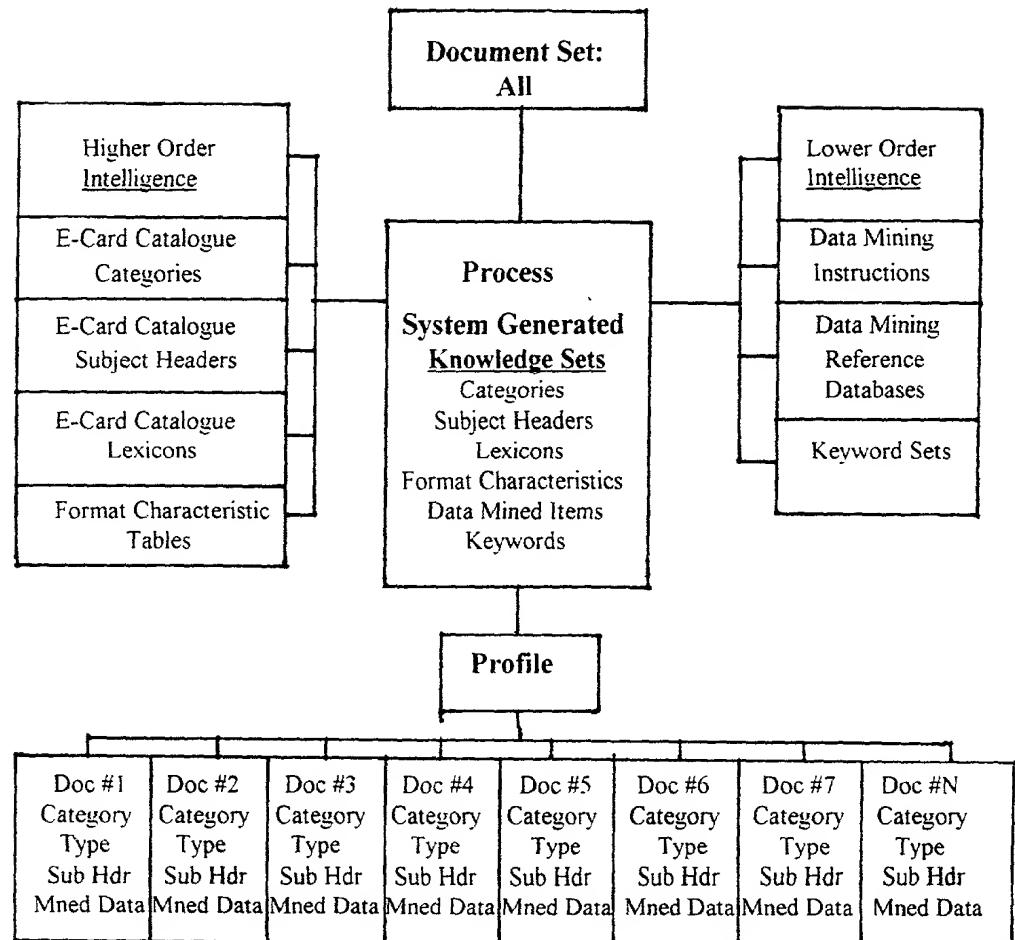
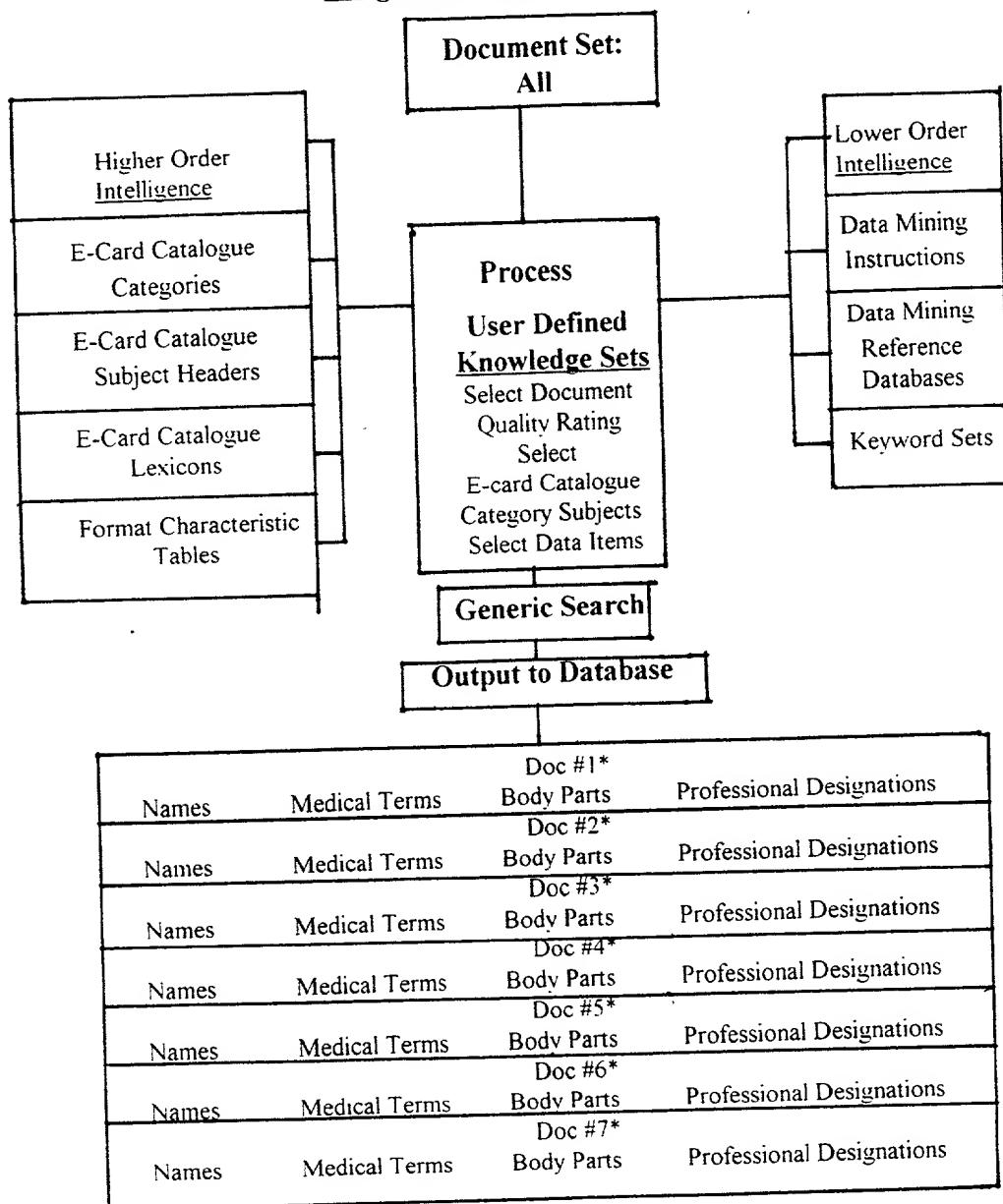


Figure 9

Diagram: Generic Searching



* as an example

Figure 10

Diagram: Data Mining

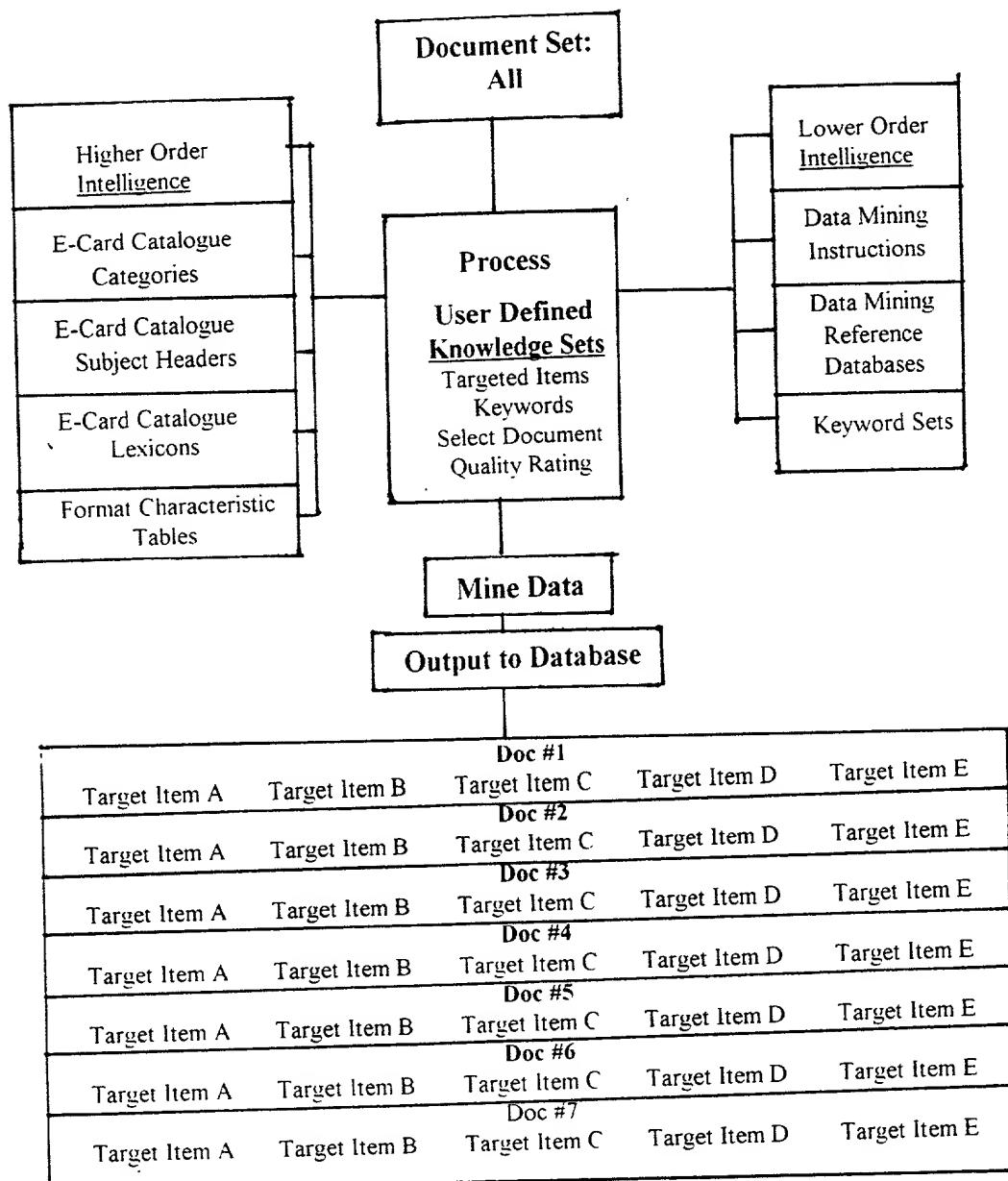
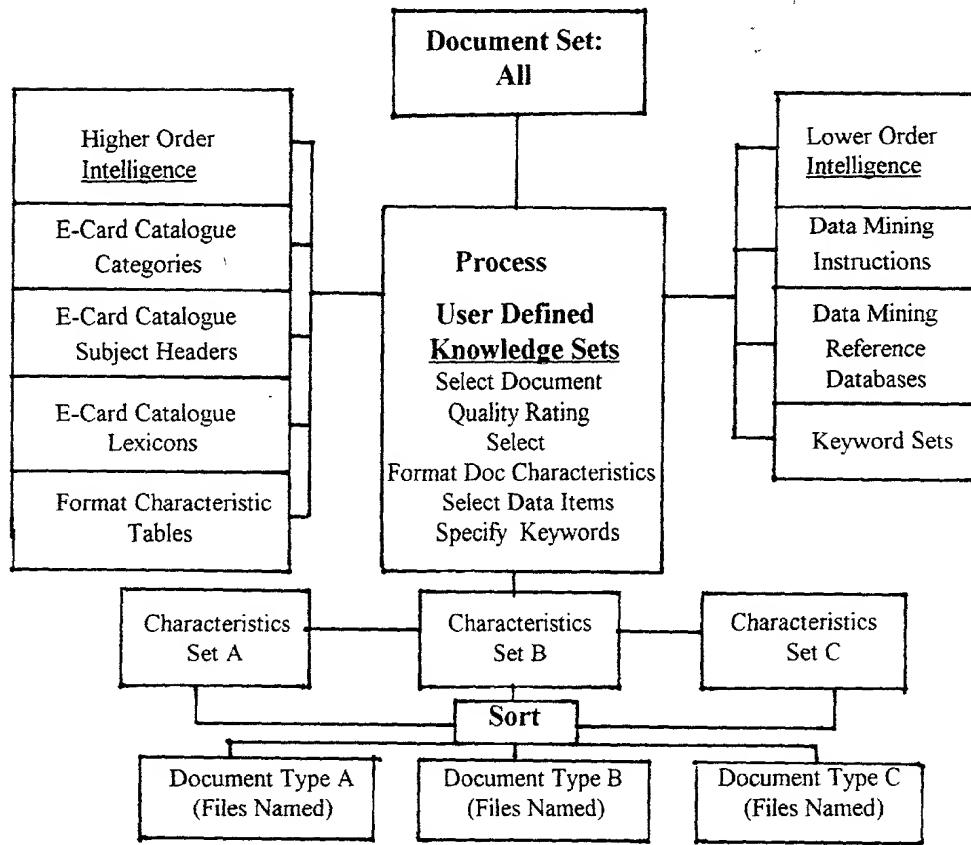


Figure 11

Diagram: Non Standard Forms Processing



Mine Data Appropriate to Document Type

| Document Type A (File Name) | Document Type B (File Name) | Document Type C (File Name) |
|--------------------------------|--------------------------------|--------------------------------|
| Mined Data (File Name) | Mined Data (File Name) | Mined Data (File Name) |
| Mined Data (File Name) | Mined Data (File Name) | Mined Data (File Name) |
| Mined Data (File Name) | Mined Data (File Name) | Mined Data (File Name) |
| Mined Data | Mined Data | Mined Data |

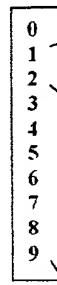
Output to Database

Figure 12

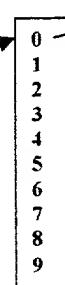
Word list and index table
 <word> <index#>

| | |
|--------------|------|
| alpha | 1004 |
| ... | |
| animal | 1027 |
| ... | |
| astrophysics | 2070 |
| ... | |
| botany | 2903 |
| ... | |
| zoology | 9900 |

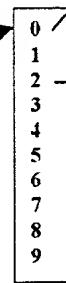
Thousands' Group



Hundreds' Grouping



Ten's Grouping



Unit Grouping
 Actual Unit Bit Array

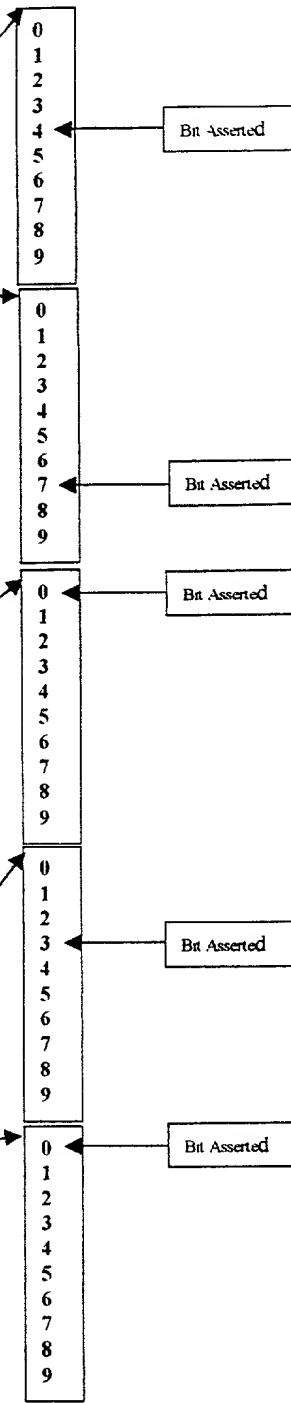


Figure 13

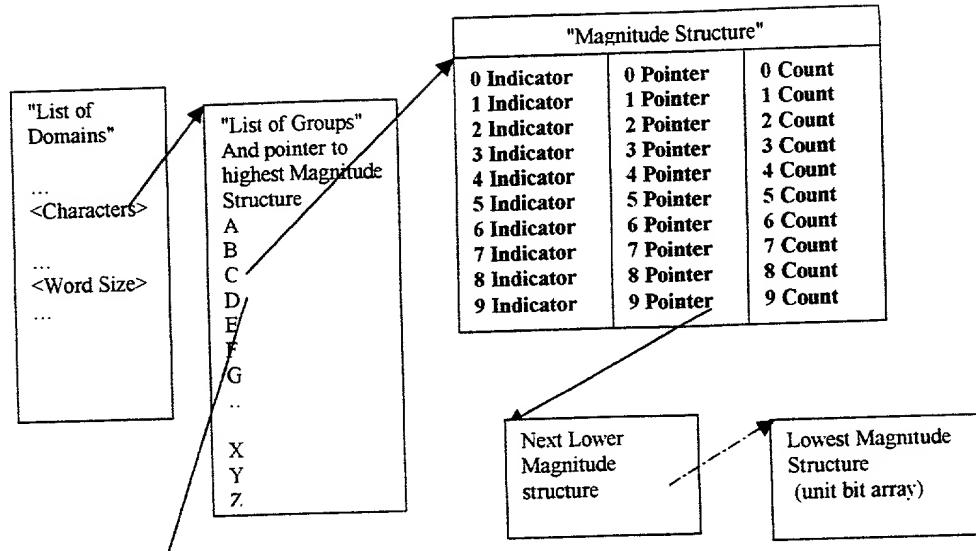


Figure 2a Actual Memory Mapping for the Full Array

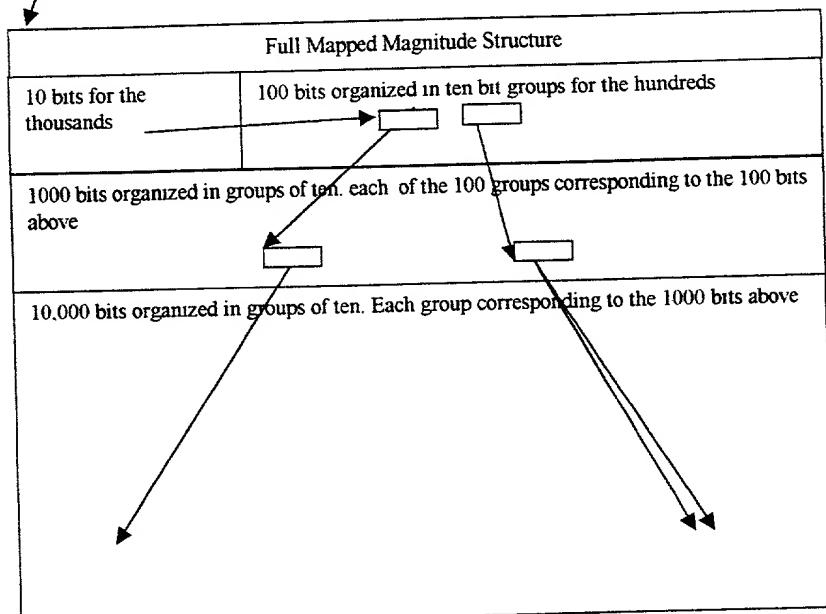


Figure 14

